Panta rhei, measurement and discovery of change in financial markets

Zamojski, M.J.

2017

document version
Publisher's PDF, also known as Version of record

Chapter 4

Filtering With Confidence: In-sample Confidence Bands For GARCH Filters

4.1. Introduction

There is vast empirical evidence that for many economic variables conditional variances and covariances change over time. Given the importance of heteroscedasticity in finance and macroeconomics[1] it is not surprising that estimation of the time-varying volatility has attracted substantial attention in the literature. As any time-varying parameter, volatility can be modelled both with observation- and parameter-driven models.[2] The observation-driven approach is most notably represented by the family of generalised autoregressive conditional heteroscedastic (GARCH) models originated by Engle (1982) and Bollerslev (1986). Stochastic volatility models are a typical example of the parameter-driven class.[3] Additionally, volatility is also estimated with realised measures which are model-free and specific to the literature on heteroscedasticity.[4] Of

[1] In macroeconomics, failure to account for time-varying volatility may lead to incorrect inference and inefficient estimates of parameters of interest (Engle, 2001; Hamilton, 2010). Fernández-Villaverde and Rubio-Ramírez (2013) further argue that many extensions in the dynamic stochastic general equilibrium (DSGE) model can be achieved by allowing for heteroscedasticity.

[2] A detailed discussion of the two approaches can be found in Cox (1981).

[3] Shephard (2005) gives a review of the subject.

[4] See, e.g., Andersen, Bollerslev, Diebold, and Labys (2001); Barndorff-Nielsen and Shephard (2002); Andersen, Bollerslev, Diebold, and Labys (2003); Corsi (2009).

the three methods, the GARCH family is the most widely used in practice due to its computational simplicity and data availability.

All three methods are routinely applied to the same data and produce estimates of static parameters, confidence intervals for these parameters, and estimates of the volatility path. In addition, the (parameter-driven) stochastic volatility and realised volatility models both produce confidence bands for the time-varying volatility (Barndorff-Nielsen and Shephard, 2002; Durbin and Koopman, 2012, §14.5), even in the absence of parameter uncertainty. By contrast, the volatility paths estimated with the GARCH models are reported without such confidence bands. Unless a researcher truly believes that a chosen GARCH model is the data-generating process and there is no parameter uncertainty, the lack of precision measures around the path of there time-varying parameter puts the GARCH family at a disadvantage.[5] This chapter introduces a novel bootstrap procedure which allows the econometrician to draw confidence bands around volatility paths estimated by GARCH models. I accomplish this by treating these models as misspecified filters rather than as true data-generating processes. I am able to account for both filtering[6] and parameter uncertainty.

The lack of confidence bands for filtered paths of time-varying parameters is not unique to GARCH models. In fact, most—if not all—observation-driven methods suffer from this shortcoming. Among these are the autoregressive conditional duration (ACD) model of Engle and Russell (1998), the dynamic conditional correlation (DCC) model of Engle (2002), the dynamic copula models of Patton (2006), or the class of generalised autoregressive score models of Creal, Koopman, and Lucas (2013) and Harvey (2013), to

[5] For instance, for approximately a quarter after the default of Lehman Brothers in September 2008, annualised volatility of S&P 500 returns had been oscillating around the 70% level based on estimates of many GARCH models, including: GARCH(1,1) of Engle (1982); Bollerslev (1986), GJR-GARCH(1,1) of Glosten, Jagannathan, and Runkle (1993b), t-GAS(1,1) of Creal, Koopman, and Lucas (2011), etc.; see NYU Stern's Volatility Lab page for a figure with these results. It is unknown, however, how likely it is that the true unobserved volatility was indeed at 70% and not, e.g., at 50% or 100% at any point in time in that period. The S&P 500 case is explored in Section 4.7. In particular, I show estimated volatility path with the LITE bands in Figure 4.1.

[6] Filtering and model uncertainty are results of different ways in which time-varying models can be misspecified. A brief explanation is provided later, for a detailed discussion see Blasques, Koopman, Łasak, and Lucas (2016).

name just a few. In addition, the generalised autoregressive method of moments, which is developed in Chapter 3, is also observation-driven and suffers from the same defect. The procedure described in this chapter can be potentially applied to all these models. Thus, the GARCH family is just one possible illustration of a wider class of models for which my bootstrap procedure can be implemented.

With that in mind, consider a typical representative of the GARCH family, the GARCH(1,1) model:

$$y_t = \varepsilon_t, \tag{4.1}$$

$$\varepsilon_t = \eta_t \sqrt{\sigma_t^2}, \tag{4.2}$$

$$\sigma_{t+1}^2 = \omega + \alpha y_t^2 + \beta \sigma_t^2, \tag{4.3}$$

where y_t denotes the observed returns and the innovations, η_t, are drawn independently from a distribution with zero mean and unit variance. For clarity of exposition, I continue with this example in the remainder of the chapter.

Blasques, Koopman, Łasak, and Lucas (2016) define three different sources of uncertainty which can play a role in modelling time-varying parameters: model, parameter, and filtering uncertainty. First, model uncertainty reflects the fact that both the true distribution of η_t and the true observation eq. (4.1) are not known. I do not focus on this type of uncertainty. Parameter uncertainty refers to the fact that the static parameters in the model (here ω, α, and β) are replaced by their maximum likelihood estimators. Finally, filtering uncertainty stems from possible misspecification in the transition eq. (4.3) for σ_t^2, for instance, if a GARCH model is used to estimate a volatility which, in fact, follows a stochastic volatility process. Note that Nelson (1992) shows that under specific conditions GARCH(1,1) produces volatility paths that converge to the true path as the sampling frequency is increased, i.e., as we go from monthly, through daily, to hourly (and so forth) data.

Assume, for now, that there is no parameter uncertainty, i.e. that 'pseudo-true'

values of the static parameters are known.[7] If the GARCH(1,1) model is assumed to be the true data-generating process, then as is evident from eqs. (4.1) to (4.3), η_t is the only source of randomness. Thus, given an observed time-series of returns and in the absence of uncertainty about the *static* parameters, there is also no uncertainty about the *dynamic* parameters—there is no need for confidence bands. Conditional on the data, the GARCH(1,1) model gives a single, deterministic path. This path coincides with the true path only if the GARCH model is indeed the data-generating process. At first sight, there is also no way of generating confidence bands around such deterministic paths. For example, simulating from the GARCH(1,1) data-generating process yields new time-series and new paths of volatility. However, these paths can lie nowhere near the path estimated for the empirical data. By contrast, the bootstrap procedure proposed here—which I call the Local In TimE (LITE) bootstrap—retains volatility patterns which are observed in the data.

The first contribution of this chapter is the design of the bootstrapping approach for GARCH filters under misspecification. In particular, I obtain confidence bands which account for the filtering uncertainty. The essence of the approach is to draw bootstrap samples conditional on the observed (filtered) volatility path. In fact, I do not use estimates of the static parameters in constructing bootstrap time-series as I do not sample from the GARCH(1,1) data-generating process. More precisely, I replace the typical *i.i.d.* residual bootstrap used so far in the literature (Hall and Yao, 2003; Gonçalves and Kilian, 2004; Pascual, Romo, and Ruiz, 2006):

$$y_t^* = \hat{\eta}_k^* \sqrt{\hat{\omega} + \hat{\alpha} y_{t-1}^{*2} + \hat{\beta} \sigma_{t-1}^{*2}}, \tag{4.4}$$

by:

$$y_t^* = \hat{\eta}_k^* \sqrt{\hat{\sigma}_t^2}, \tag{4.5}$$

where subscript k on $\hat{\eta}_k^*$ means that the raw (normalised) residuals are resampled. LITE bootstrap samples are obtained without using the transition eq. (4.3) implied by the

[7] Blasques, Koopman, and Lucas (2014) define the pseudo-true parameter as a unique maximiser of the log likelihood in the limiting case, i.e., when the number of observations tends to infinity.

filter. Instead, I directly re-scale the resampled residuals using the fixed volatility path, $\hat{\sigma}_t^2$, estimated from the empirical data. In this way, paths in the bootstrap world preserve volatility patterns which are observed in the data. Figure 4.1 compares the two approaches. In Panel A, the real world data is generated by a GARCH model, i.e., time t innovation determines the following period's volatility. The aim of the bootstrap is to replicate the entire data-generating process and estimation (solid lines) in the 'bootstrap world'. The typical implementation replaces the unknown static parameters with their estimates (dashed arrow). This exemplifies the plug-in principle (see, e.g. Efron, 2003). In Panel B, volatility is not generated by a GARCH model, and I replace the *i.i.d.* bootstrap with the LITE procedure. Just like in the *i.i.d.* bootstrap, estimation steps (arrows) in Panel B are the same in both worlds. The main difference between the two approaches lies in which point estimates are retained from the real world to construct samples in the bootstrap world. The *i.i.d.* bootstrap uses raw residuals and estimates of the static parameters, discarding the empirical path of volatility. In contrast, I propose to use the estimates of the dynamic parameter along with the residuals.

Due to the plug-in principle, the LITE bootstrap accounts for the parameter uncertainty as well as the filtering uncertainty discussed before. In comparison, the method proposed by Blasques, Koopman, Łasak, and Lucas (2016) accounts only for the parameter uncertainty as they explicitly assume that the filter is the true model. Thus, LITE seems to be more relevant for financial applications where the available time-series are both long and can be sampled at very high frequencies decreasing the impact of parameter uncertainty.

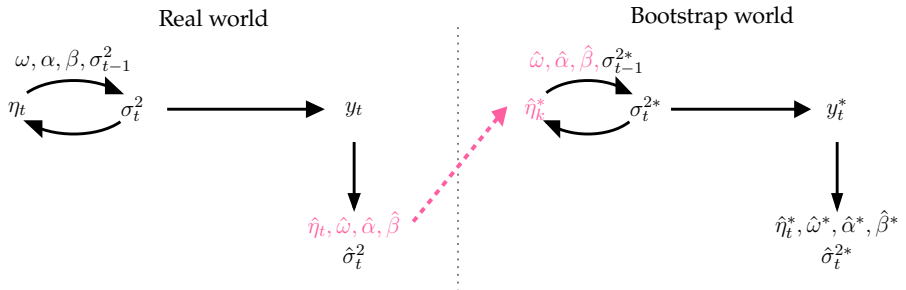
I show that simply taking the estimated path of volatility and implementing an *i.i.d.* resampling (Efron, 1979; Freedman, 1984) of residuals produces poor results. The resulting bootstrap paths are over-smoothed, inconsistent, and cannot be used to create confidence bands with good coverage properties. The second innovation of this chapter is, therefore, the way in which I set up the resampling within the bootstrap procedure. To overcome the above difficulty, I consider residuals to be valid replicates only within

Figure 4.1

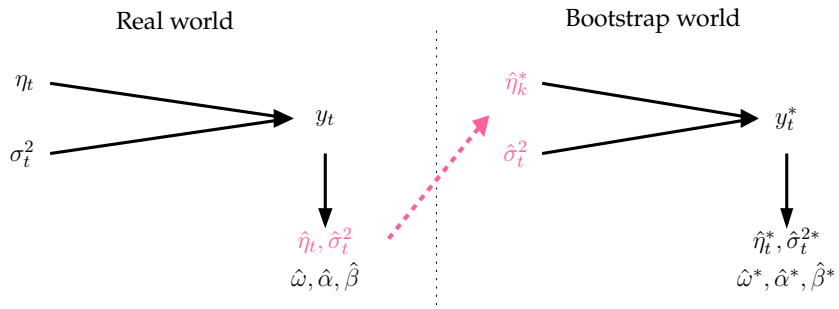
Point of novelty

This figure illustrates the difference between a typical *i.i.d.* residual bootstrap and the LITE bootstrap. Vertical arrows correspond to estimation. Point estimates taken from the real world to construct samples in the bootstrap world are coloured red and presented in the first line.

Panel A: GARCH is the DGP, *i.i.d.* residual bootstrap



Panel A: GARCH is not the DGP (just a filter), LITE residual bootstrap



a local window of time. This ensures that values of not only true volatilities but also local measurement errors are close to one another. Matching misfit both in value and in sign turns out to be of great importance in ensuring good performance of the bootstrap. The resampling proceeds in a manner resembling a moving-window which motivates the 'Local In TimE' name of the entire procedure. The length of the window depends on the persistence of the underlying process. Efron's, *i.i.d.* rule is a special case of the LITE bootstrap.

I evaluate the performance of the LITE bootstrap using a range of simulation experiments. I find that the average coverage of LITE confidence bands is close to the nominal level set by the econometrician. Moreover, LITE confidence bands behave re-

markably similar to those which could have been obtained in Monte Carlo experiments, i.e. when both the true path of volatility and the true conditional densities are known. In addition to generating confidence intervals, LITE can also act as a smoother and mitigate the impact of sampling error.[8] Based on simulations, the paths obtained with the LITE smoother are characterised by significantly lower root mean square errors than the GARCH paths. LITE volatility paths appear to inherit consistency properties (in the sense of Nelson, 1992) of the underlying GARCH filter. Simulation evidence suggests that as sampling frequency is increased, both GARCH and LITE paths seem to converge to the true latent volatility path and the LITE confidence bands shrink. Description of the asymptotic behaviour of the LITE confidence bands is left for future research. Finally, the LITE bootstrap is easily implementable and does not impose a significantly higher computational burden.

Volatility modelling has been most popular in finance where the trade-off between risk and return is emphasised in theories of asset pricing. Estimates of the latent volatility and its forecasts are key inputs in asset valuation and risk management tools. Standard volatility models can be used as well to measure systemic risk (Engle and Siriwardane, 2015), bringing volatility modelling to the forefront of current policy discussion. Additional information provided by the LITE confidence intervals is, thus, meaningful to both investors and policy-makers since the LITE procedure automatically constructs a confidence interval for the one-period-ahead forecast provided by the GARCH model. This allows us to compute similar confidence intervals for widely used risk measures like value-at-risk (VaR) or expected shortfall. Furthermore, I observe that the width of the confidence intervals changes over time even after conditioning on the estimated volatility path. This can be taken advantage of in many investment strategies, e.g., risk targeting or risk parity.

This chapter contributes to two strands of the literature. First, this chapter adds to the vast literature on observation-driven models and allows, for the first time, to plot

[8]Sampling error causes a well-established attenuation bias, where estimated volatility path based on observed data is characterised by lower persistence than in the underlying data-generating process.

confidence bands around time-varying parameters in a way that accounts for both the filtering and the parameter uncertainty. Although bootstrapping techniques have been previously used for GARCH models, the focus thus far has been on obtaining confidence intervals only for the static parameters see e.g., Hall and Yao (2003); Gonçalves and Kilian (2004); Pascual, Romo, and Ruiz (2006). As a by-product of the procedure, LITE produces confidence intervals for these parameters as well, but I propose to use them mainly to gauge the level of uncertainty due to filtering.

Secondly, the LITE resampler adds to the general bootstrap literature. The resampler generates pseudo time-series which imitate the true—unobserved—dependence structure along with conditional distributions at all times. Although fundamentally different, LITE is close in spirit to the 'local bootstrap' of Paparoditis and Politis (2000, 2002) where locality is defined in parameter space,[9] not in time. Their 'local bootstrap' does not capture potential persistence of misfit yielding bootstrap samples with either too fat or too thin tails. Locality in the time dimension is a stricter condition as it also accounts for persistence of misfit.

The remainder of this chapter is organised as follows. Section 4.2 describes the setting and the GARCH(1,1) filter I focus on. In Section 4.3, I describe LITE bootstrap procedure. I elaborate on the moving-window resampler in Section 4.4. In Section 4.5, I discuss how performance of the confidence bands is measured and evaluated in this chapter. Section 4.6 shows how to operationalise the choice of the window length (bandwidth). I implement the procedure in an empirical study of S&P 500 returns in Section 4.7. In Section 4.7, I look at coverage of the LITE confidence intervals in a series of simulation experiments. Finally, Section 4.8 concludes.

[9] This criterion may not be able to ensure validity of replicates if the filter is misspecified and/or the filtered path lags behind the true path. For instance, estimated volatility in two periods can indeed be close. But, if the volatility in one of the periods was increasing while it was decreasing in the other, the true volatility is respectively under- and over-estimated by a lagging filter. Drawing replicates indiscriminately from both periods yields conditional distributions in the bootstrap samples with either too fat or too thin tails.

4.2. The GARCH filter

Consider a stationary sequence $Y = \{y_i : 0 \leq i \leq T\}$ where T is the size of the observed sample. Let σ_t^2, $\hat{\sigma}_t^2$, $\hat{\sigma}_t^{2(*,b)}$ denote the true value of the latent volatility at time $t = 1, \ldots, T$, the estimated value, and the estimate based on the b^{th} bootstrap time-series, respectively. I use a short-hand notation, $\{\sigma_t^2\}$, to denote the full, realised path of the time-varying volatility $\{\sigma_1^2, \sigma_2^2 \ldots \sigma_T^2\}$. By analogy, $\{\hat{\sigma}_t^2\}$ and $\{\hat{\sigma}_t^{2(*,b)}\}$ denote fitted paths based on the original sample and the b^{th} bootstrap sample.

The data-generating process considered here is as follows:

$$\begin{aligned} y_t &= \varepsilon_t, \\ \varepsilon_t &= \eta_t \sqrt{\sigma_t^2}, \qquad & \eta_t \sim \mathrm{N}(0,1), \\ \sigma_t^2 &= M\left(\sigma_1^2, \ldots, \sigma_{t-1}^2, y_1, \ldots, y_{t-1}\right), \end{aligned} \tag{4.1}$$

where the true conditional variance σ_t^2 follows a process assumed to be unknown to the econometrician. Although the true dependence model, M, is not known it is assumed to produce sufficiently smooth paths of volatility such that it can be *approximated* with the GARCH(1,1) filter (updating rule):

$$\hat{\sigma}_{t+1}^2\left(\hat{\sigma}_t^2, y_t, \omega, \beta, \alpha\right) = \omega + \beta\hat{\sigma}_t^2 + \alpha y_t^2, \tag{4.2}$$

where α, β, and ω need to be estimated.

One example of a dependence model which can be successfully approximated with the GARCH(1,1) filter is the stochastic volatility process. For such data-generating processes, Nelson (1992) shows that under certain regularity conditions GARCH(1,1) produces a consistent estimate, $\{\hat{\sigma}_t^2\}$, of the unknown path $\{\sigma_t^2\}$. Among the conditions is smoothness of the volatility path and. In particular, presence of jumps (or other forms of discontinuity) in the volatility process usually breaks down the consistency immediately around the point of the jump. However, even in these cases observation-driven filters can be a useful tool in extracting paths of time-varying parameters.

4.3. Local in time bootstrap

In this section, I describe the lite bootstrap procedure. For clarity of exposition, I refer to the path of the time-varying volatility estimated from the empirical data, $\{\hat{\sigma}_t^2\}$, as the 'GARCH path'. Paths obtained in the bootstrap procedure, $\{\hat{\sigma}_t^{2(*,b)}\}$, are collectively called 'bootstrap paths'.

The lite procedure takes the 'GARCH path' ($\hat{\sigma}_t^2$) and GARCH residuals as inputs. The role of the algorithm is to create valid bootstrap samples, meaning pseudo time-series which sufficiently recreate the true—unobserved—dependence structure in the data. In this way, the bootstrap world poses the same challenges for the filter as the real world.

The lite bootstrap algorithm takes the following five steps:

Step 1: Estimate $\{\hat{\sigma}_t^2\}$ with the GARCH(1,1) filter and obtain the corresponding series of fitted residuals $\{\hat{\varepsilon}_t\}$. Subsequently, create a series of raw residuals, $\{\hat{\eta}_t\}$, by adjusting each $\hat{\varepsilon}_t$ for scale, such that $\hat{\eta}_t = (\hat{\sigma}_t)^{-1}\hat{\varepsilon}_t$.[10]

Step 2: Generate B bootstrap series of raw residuals, $\{\hat{\eta}_t{}^{(*,b)}\}$,by sampling with replacement from raw residuals using the LITE resampler which is described in detail in Section 4.4.[11]

Step 3: Re-scale the raw bootstrap residuals with the GARCH path, $\{\hat{\sigma}_t^2\}$, from Step 1 to obtain B bootstrap samples with a typical element, $y_t{}^{(*,b)} = \varepsilon_t{}^{(*,b)} = \hat{\sigma}_t\hat{\eta}_t{}^{(*,b)}$.

Step 4: For each bootstrap sample, $\{y_t{}^{(*,b)}\}$, estimate a volatility path, $\{\hat{\sigma}_t^{2(*,b)}\}$, with the GARCH(1,1) filter.

Step 5: The empirical distribution function of the time-varying volatility at time t is

[10] In the case of volatility modelling, the raw residuals can be further demeaned but should not be standardised, see discussion in Section 4.4.

[11] At this stage, assume the LITE resampler uses the standard Efron's rule (*i.i.d.* resampling), but applies it in a moving-window fashion.

defined as:

$$F_{B,t}(x) = \Pr_{B,t}\left(\sigma_t^2 \leq x\right) = \frac{1}{B+1}\sum_{i=1}^{B+1} \mathbb{1}\left(\sigma_t^{2(*,b)} \leq x\right), \tag{4.1}$$

where $\mathbb{1}(\cdot)$ is an indicator function returning 1 if the condition in parentheses is true. Confidence intervals for $\hat{\sigma}_t^2$ are inferred from the empirical distribution function and defined as:

$$\left[F_{B,t}^{-1}\left(\frac{a}{2}\right), F_{B,t}^{-1}\left(1-\frac{a}{2}\right)\right].$$

If B is chosen such that $\frac{a}{2}(B+1)$ is an integer, this simplifies to selecting the $\frac{a}{2}(B+1)^{\text{th}}$ and $(1-\frac{a}{2})(B+1)^{\text{th}}$ value of a sorted series $\{\hat{\sigma}_t^2, \hat{\sigma}_t^{2(*,1)}, \ldots, \hat{\sigma}_t^{2(*,B)}\}$.[12] The bootstrapped confidence intervals can be computed to arbitrary precision by increasing B.

In contrast to previous literature (e.g., Hall and Yao, 2003; Gonçalves and Kilian, 2004; Pascual, Romo, and Ruiz, 2006), the LITE bootstrap does not require simulating new volatility paths from the transition equation implied by the filter. Instead, in Step 3, I propose to rescale resampled residuals in the bootstrap world with the volatility path that was filtered out from the empirical data. As such, estimates of the static parameters governing the updating rule are not directly used in the creation of bootstrap samples. It is precisely the omission of this step that allows LITE to create pseudo time-series for which newly filtered paths are relatively close to the considered 'GARCH path' at all times and consistently across bootstrap samples. This is the first contribution of the chapter.

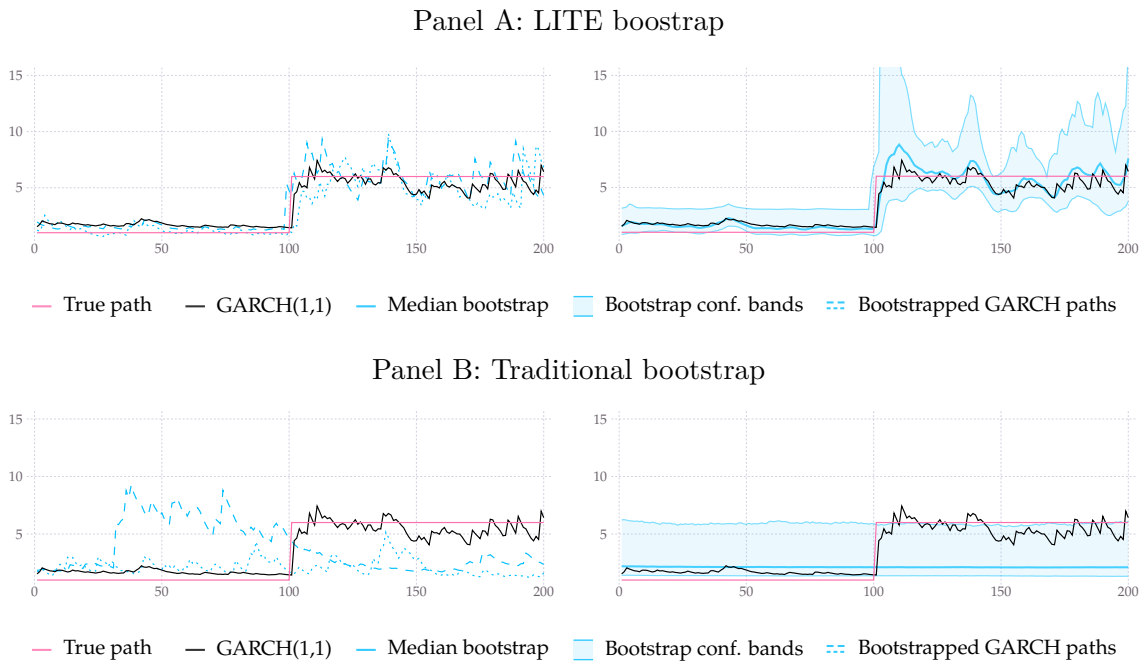
To develop intuition on how does the LITE procedure achieve this, consider a motivating example. In Figure 4.1, I look at a simple scenario in which the volatility process exhibits a structural break in the middle of the visible sample. In Panel A, I use the LITE procedure while Panel B contains results obtained with a traditional approach. In

[12] Assuming filtered paths are stored as row vectors, vertically stack bootstrap paths $\{\hat{\sigma}_t^{2(*,b)}\}$ and the GARCH path $\{\hat{\sigma}_t^2\}$ from Step 1 together to form a $(B+1)\times T$ matrix. A series $\{\hat{\sigma}_k^2, \hat{\sigma}_k^{2(*,1)}, \ldots, \hat{\sigma}_k^{2(*,B)}\}$ corresponds to the k^{th} column of this matrix and is the sole input in producing a confidence interval for that period of time.

Figure 4.1

Motivation for the LITE bootstrap

This figure illustrates differences between the traditional and the LITE bootstrap procedures. I consider a simple scenario where the volatility regime changes from low to high in the middle of the visible sample. I plot the true volatility path in pink and the filtered GARCH(1,1) path in black. The initial adjustment period of 50 observations is removed. In each panel, the plots on the left-hand side contain two bootstrapped volatility paths. These paths are drawn as dashed lines. The plots on the right-hand side contain 90% confidence bands which can be obtained with the respective bootstrap procedure. Panel A contains results for the LITE bootstrap. In Panel B, I plot results which would be obtained with a traditional bootstrap procedure. Here, the GARCH(1,1) transition equation is used to draw *new* volatility paths while generating bootstrap samples.



both cases, the plots on the left-hand side contain the true volatility path, the estimated 'GARCH path', as well as two 'bootstrap paths', i.e., volatility paths filtered from bootstrap samples. Plots on the right-hand side show the resulting 90% confidence bands obtained based on 3999 bootstrap replications. As the LITE constructs bootstrap samples based on the 'GARCH path', the 'bootstrap paths' share similar dynamic properties with it. Furthermore, the 'bootstrap paths' preserve temporal patterns of volatility that are seen in the original data. Confidence bands created with the LITE reflect this and tightly follow the 'GARCH path.' The bands appear to provide good coverage as well.

Due to this, they are meaningful and can be used for hypothesis testing.

Panel B contains results obtained with a traditional bootstrap procedure. In the traditional approaches, bootstrap-true paths[13] are generated recursively based on the filter's transition equation. These paths will have the same dynamic properties, but do not preserve temporal patterns. As a result, the 'bootstrap paths' can be nowhere near the 'GARCH path'. For instance, the left-hand side plot of Panel B contains one filtered path which indicates there was hardly any period of high volatility in its bootstrap sample. Although the second bootstrap sample did include a prolonged period of high volatility, it occurred too early. It is clear from the right-hand side plot of Panel B that confidence bands for volatility paths obtained with the traditional bootstrap procedures are not informative at any particular point in time. In fact, the true volatility is outside of the 90% confidence band almost throughout the whole sample in this example. The bands become even less meaningful if the procedure is applied to empirical data that includes relatively short, but economically meaningful, periods of extremely high volatility. In such a setting, the confidence bands would label crisis periods, which are usually characterised by an increase in volatility, as outliers in their entirety. This is not surprising. The traditional approaches were never meant to provide bands around a path of the time-varying volatility. Instead, the focus of this strand of literature is on measuring uncertainty around estimates of the static parameters in the filter under the assumption that the filter is also a good description of the data-generating process. As these bootstrap procedures create samples unconditionally from the model, the resulting confidence bands are reflective of, at most, the unconditional value of the volatility. In comparison, the LITE bootstrap creates pseudo time-series conditional on the observed data.

The (percentile) confidence intervals computed in Step 5 are invariant under monotone transformations. Thus, LITE can be used to produce bands for volatility even if the filter is specified in terms of variances or log-volatilities. Due to this property, it is

[13] The true values of parameters in the bootstrap world. For LITE this would be the 'GARCH path.'

also possible to produce confidence intervals for statistics that depend on variance, like value-at-risk or expected shortfall.

4.4. LITE resampler

The second contribution of this chapter is the LITE resampler which attempts to ensure that observations in bootstrap samples are drawn from correct conditional distributions even if the 'GARCH path' locally under- or over-estimates the true conditional volatility.

Although temporal patterns of volatility in the LITE procedure are enforced through the rescaling of the raw residuals with the 'GARCH path' in Step 3, the misspecified nature of the filter creates an additional obstacle. Ideally, one would sample ${\varepsilon_t}^{(*,b)}$ from the same distributions as ε_t. In Step 3, ${\varepsilon_t}^{(*,b)}$ are constructed as the product of $\hat{\sigma}_t$ and resampled $\hat{\eta}_t$. Due to misspecification, the 'GARCH path', $\hat{\sigma}_t$, locally under- or over-estimates the true volatility which affects conditional distributions of raw residuals. The local misfit, though persistent, is both changing in time and unknown. As a result, special attention to finding valid replicates for $\hat{\eta}_t$ is required. In particular, while consecutive raw residuals are likely to be under- or over-scaled to the same degree, this cannot be guaranteed for any two randomly selected $\hat{\eta}_t$. As the nature of the resampler dictates which raw residuals are considered as valid replicates at any point in time, the quality of the obtained confidence bands will depend on the type of resampling which is used in Step 2.

The LITE resampler operates in a moving-window fashion and draws replicates for $\hat{\eta}_t$ from a 'small' neighbourhood around time t. The size of the window is determined by a bandwidth parameter, w. Let $\delta(t) \sim \Delta_t^w$ be a random offset whose distribution function assigns equal probabilities[14] to elements in a time-varying sample space:

$$\{i : \max[1, t-w] \leq i \leq \min[t+w, T], i \in \mathbb{Z}\}, \tag{4.1}$$

[14] The uniform kernel used here can be replaced by other kernels with compact support.

with $T \geq w \geq 1$. Given a b^{th} draw, $\delta_t^{(b)}$, from the distribution Δ_t^w, the raw residual $\hat{\eta}_{t+\delta_t^{(b)}}$ is drawn as a replicate for the residual at time t. Analogously $\hat{\sigma}^2_{t+\delta_t^{(b)}}$ is the corresponding variance used to normalise the replicating residual. The resampling bandwidth can be chosen to minimise squared bias of bootstrap paths w.r.t. the 'GARCH path' (see Section 4.6 for details). A bootstrap observation at time t is then obtained as:

$$y_t{}^{(*,b)} = \varepsilon_t{}^{(*,b)}, \tag{4.2}$$

$$\begin{aligned} \varepsilon_t{}^{(*,b)} &= \hat{\eta}_{t+\delta_t^{(b)}} \sqrt{\hat{\sigma}_t^2} \\ &= \frac{\hat{\varepsilon}_{t+\delta_t^{(b)}}}{\sqrt{\hat{\sigma}^2_{t+\delta_t^{(b)}}}} \sqrt{\hat{\sigma}_t^2}. \end{aligned} \tag{4.3}$$

To illustrate the rationale behind the LITE resampling procedure, I again consider a case in which the volatility process exhibits structural breaks. In Figure 4.1, the volatility switches to a high regime at time $t = 100$ and returns to the low regime at $t = 250$. I plot the true and the filtered path of the volatility in Panel A. In Panel B, I look at the degree to which the raw residuals, $\hat{\eta}_t$, are under- or over-scaled due to the misfit of the 'GARCH path'. For instance, at the beginning of the sample the volatility is estimated relatively too high. The raw residuals are over-scaled and are too small when compared with the true η_t. Thus, the difference between $|\hat{\eta}_t|$ and $|\eta_t|$ tends to be negative for $t < 100$. Similarly, in the high volatility regime, the 'GARCH path' tends to be below the true path which results in under-scaled raw residuals. The severity of this problem is the highest shortly after a structural break. Furthermore, a Monte Carlo experiment based on 10,000 replications reveals that the misfit and the resulting under- or over-scaling of raw residuals (the blue line in Panel B) is not due to sampling bias, but indeed a consequence of misspecification.

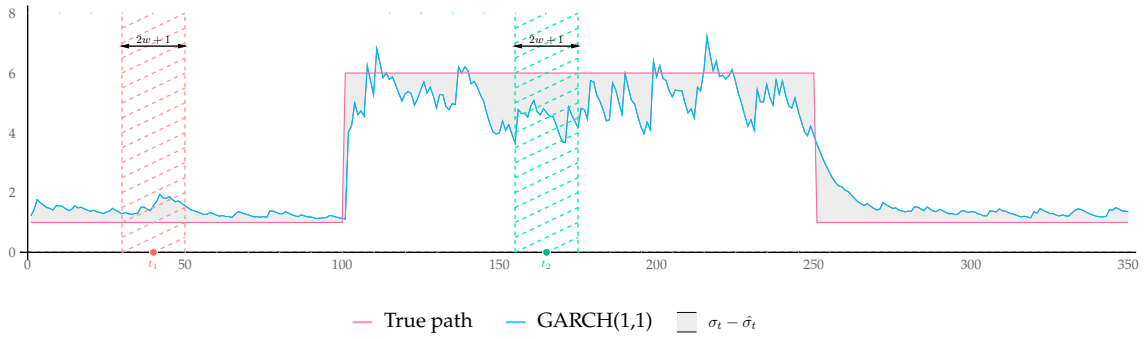
Consider the problem of selecting a valid replicate for an observation at time t_2. In this case, the volatility is estimated too low. The LITE resampler exploits the fact that, in a sufficiently small window of time around t_2, observations share three characteristics: (i) the unobserved, true volatility values are similar, (ii) the estimated values are close, and consequently (iii) the volatility estimates in this neighbourhood suffer from a similar

Figure 4.1

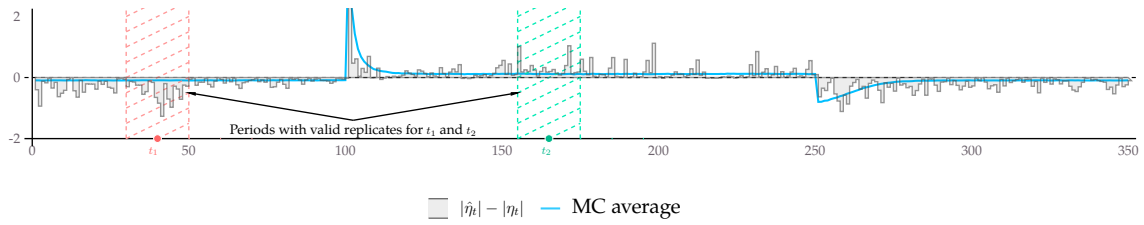
Motivation for the LITE resampler

This figure illustrates the local resampling procedure. Panel A shows a setting where there are two breaks in the true volatility path. At time $t = 100$ the volatility shifts from a low to a high regime and subsequently returns to the low regime at $t = 250$ of the visible sample period. The initial adjustment period is removed. I plot the true path of the volatility and the filtered GARCH(1,1) path. In Panel B, I plot in grey differences between absolute values of raw residuals, η_t, from the true η_t. For reference, I also provide a Monte Carlo average based on 10,000 replications in blue.
Consider two observations, at time t_1 and t_2 respectively, for which bootstrap replicates are drawn. The striped regions around points t_1 and t_2 show periods with valid replicates based on the resampling bandwidth of $w = 10$. Note that within these two periods observations share three characteristics: (i) the unobserved, true volatility values and (ii) the estimated values are close. Consequently, (iii) they are encumbered by similar levels of misfit. In particular, the volatility is overestimated around t_1 and underestimated around t_2. Raw residuals from the period around t_2 are not valid replicates for residual at t_1 as they are under-scaled and, in effect, their distribution has higher standard deviation than is required to imitate the conditional distribution around t_1.

Panel A: GARCH(1,1) volatility, $\hat{\sigma}_t$



Panel B: Deviations of absolute raw residuals, $|\hat{\eta}_t| - |\eta_t|$



misfit. It follows, that the raw residuals around t_2 can be expected to be under-scaled to a similar degree and are, therefore, valid replicates for $\hat{\eta}_{t_2}$. The larger the persistence of the local misfit, the more effectively the LITE resampler is able to account for the misfit.

In particular, if the replicate raw residual is under-scaled to the same degree as $\hat{\eta}_{t_2}$, the resulting bootstrap innovation ${\varepsilon_{t_2}}^{(*,b)}$ is drawn from the same conditional distribution as ε_t.

In contrast, raw residuals around time t_1 are not valid replicates for the observation at time t_2. This is because the volatility at t_1 is estimated too high and the corresponding raw residuals are over-scaled. A bootstrap innovation ${\varepsilon_{t_2}}^{(*,b)}$ constructed by multiplying an over-scaled raw residual by an underestimated value of volatility would then be in fact drawn from a distribution whose variance is much smaller than σ_t^2.

For 'small enough' values of the bandwidth parameter, the LITE procedure effectively replaces ε_t with a different draw from a conditional distribution which is as close to the true one as possible. In addition to the confidence intervals, it is then possible to obtain a smoothed version of $\{\hat{\sigma}_t^2\}$ by computing the average or the median bootstrap estimate. Simulations in the empirical part of the chapter suggest that the smoothed series are characterised by lower average root mean square error than GARCH paths from Step 1.

If bandwidth is set to be equal to the number of available observations ($w = T$), LITE recovers the *i.i.d.* bootstrap of Efron (1979), Freedman (1984), and Efron and Tibshirani (1986). However, based on the simulation evidence I present in Section 4.6, it appears it it is never optimal to choose the bandwidth this large.[15] In practice, w rarely exceeds 1% of the sample size if chosen according to a rule I propose in Section 4.6.

REMARK 4.1. Note that bootstrap procedures which do not employ a moving-window resampler are unlikely to consistently select valid replicates for misspecified filters. For instance, the local bootstrap of Paparoditis and Politis (2002) considers a residual at time k to be a potential draw for time t if $|f_t - f_k| < \epsilon$ for some $\epsilon \to 0$ even if $|t-k|$ is large.

[15] In fact, I conjecture that that as $T \to \infty$, I require that $w \to \infty$ in a way that ensures that $w/T \to 0$. In other words, as the number of observations increases the *fraction* of observations which are considered to be valid replicates tends to zero, but in such a way the *number* of valid observations goes to infinity. The increase in the number of observations here should be understood in the fill-in asymptotics sense, i.e., it is obtained by through an increase in the sampling frequency. It could, for instance, mean switching from monthly to daily, or from daily to hourly data spanning the same period of time. This is a familiar condition which is also present in block bootstrap procedures (Carlstein, 1986; Kunsch, 1989).

In contrast to LITE , this procedure is local in the parameter space and may misidentify the set of valid replicates. To see why, consider again the example in Figure 4.1. In the extreme case, although the estimates of volatility in the periods shortly after each structural break are very close, the corresponding raw residuals should not be considered valid replicates. This is because values of misfit associated with the same value of the estimated volatility are significantly different in these two periods of time. Raw residuals after the volatility increases are severely under-scaled and significantly over-scaled after volatility decreases. Defining locality based on distance in parameter space does not, in principle, ensure that misfit due to misspecification is accounted for. Conversely, if $\{\sigma_t^2\}$ was generated by a sufficiently persistent process for which observation-driven filters can be expected to perform well, defining a neighbourhood for residuals in the time dimension implies that the associated values of the filtered variance are also relatively close.

REMARK 4.2. In the bootstrap literature (see, e.g., Hall and Yao, 2003) it is common to re-centre and further rescale raw residuals in order to correct for over-fitting and for the fact that residuals may not necessarily sum to zero in non-linear models. If the filter is misspecified, however, normalising raw residuals can adversely affect performance of the bootstrap. Let $\tilde{\eta}_t = {}^{\hat{\eta}_t - \bar{\hat{\eta}}_t}/_{s(\hat{\eta}_t)}$ be a standardised raw residual. $\bar{\hat{\eta}}_t$ and $s(\hat{\eta}_t)$ are, respectively, the sample average and standard deviation of the raw residuals. Bootstrap replicates would then be defined as $\tilde{\varepsilon}^{(*,b)}{}_t = \tilde{\eta}_{t+\delta_t^{(b)}} \hat{\sigma}_t$. Consider the familiar setting where at $t = t_l$ there is a shift from a low to a high volatility regime. Given the misspecification, the filtered variance path does not adjust to the new setting until time $t = t_h$ (where $t_h > t_l$). For simplicity, assume that outside of the transition period the 'GARCH path' fits perfectly, i.e. $\forall t \notin [t_l, t_h] : \hat{\sigma}_t^2 \equiv \sigma_t^2$. This is a more stylised scenario than the one presented in Panel A of Figure 4.1. Then, the full series of raw residuals, $\{\hat{\eta}_t\}$, are drawn from a mixture of normals, with fatter tails induced by under-scaling during the adjustment window and, consequently, standard deviation above 1. Without standardising, all raw residuals outside of the transitional period would be perfect replicates for one another. However, after standardising the whole series of

raw residuals, the correctly scaled raw residuals would be scaled down. As a result, bootstrap innovations would be drawn from conditional distributions with too small variances even if a moving-window resampler was used. Thus, standardised residuals are no longer valid replicates for innovations in neighbouring time periods. In the presence of misspecification, confidence intervals constructed based on the standardised series may exhibit worse coverage properties. In real data settings, this issue especially degrades coverage of the estimated confidence bands in low volatility states.

REMARK 4.3. The LITE resampling procedure uses single raw residuals. Alternatively, it is possible to use larger blocks of observations to further recreate the dependence structure in the bootstrap world. In this case the random offset $\delta(t)$ is drawn from $\Delta_t^{w,l}$ where l is the block size and $T \geq w \geq l \geq 1$. Bootstrap residuals are drawn at times $t = \{1 + 0l, 1 + 1l, \ldots, kl\}$, where k is the number of blocks needed to match the sample length (i.e., $T = kl$).[16] If $w = T$, the procedure collapses to the overlapping block bootstrap of Kunsch (1989).

4.5. Performance evaluation

Performance of the LITE confidence bands can be assessed by comparing the actual coverage with the chosen nominal level. In this chapter, I adopt the following notion of coverage.

Definition 4.1 (Average path coverage)**.** *Let $CI_{lo}(\hat{\sigma}_t^2)$ and $CI_{hi}(\hat{\sigma}_t^2)$ denote the lower and upper bounds of the bootstrap confidence interval at time t. Then, average path coverage for all times in $\mathscr{T}$ is defined as:*

$$C(\mathscr{T}) = \frac{\sum_{t \in \mathscr{T}} \mathbb{1}\left[CI_{lo}(\hat{\sigma}_t^2) \leq \hat{\sigma}_t^2 \leq CI_{hi}(\hat{\sigma}_t^2)\right]}{\sum_{t \in \mathscr{T}} 1}. \tag{4.1}$$

I use Definition 4.1 in two contexts. First, if $\mathscr{T} = \{1, 2, \ldots, T\}$, the path coverage describes how many times (e.g., days) the true path of volatility was outside the confidence bands. In Section 4.7, I also investigate how average coverage depends on the

[16] Earlier, I conjecture that that $T \to \infty$, $w \to \infty$, and $w/T \to 0$. Given this and the restriction $w \geq l$, the usual condition on block size, $\frac{l}{T} \to 0$, holds in this case as well.

value of the true volatility. In this case, $\mathscr{T} = \{t : \sigma_C^{lo} \leq \sigma_t^2 \leq \sigma_C^{hi}\}$ where $[\sigma_C^{lo}, \sigma_C^{hi}]$ is a volatility bin. Description of the asymptotic behaviour of these bands is left for future work.

4.6. Feasible bandwidth selection

Quality of confidence bands obtained with LITE depends on the choice of resampling bandwidth. In this section, I propose a feasible rule of selecting the optimal value of this parameter. I also show simulation evidence which suggests that this feasible rule maximises coverage.

The resampling bandwidth (which acts as a smoothing parameter) follows as the solution to the following optimization problem:

$$\underset{w:\delta(t)\sim\Delta_t^w}{\operatorname{argmin}} \left\{ \frac{1}{T} \sum_t \left[\frac{1}{B} \sum_{i=1}^{B} \hat{\sigma}_t^{2(*,b)} - \hat{\sigma}_t^2 \right]^2 \right\}. \tag{4.1}$$

Note that eq. (4.1) minimises the average (in time) squared bias of the estimated bootstrap paths and not the average mean square error. A solution can be obtained by performing a grid search over w.

Figure 4.1 shows simulation evidence to support the selection rule in eq. (4.1). In each panel I adopt a different criterion for choosing the optimal bandwidth. Panel A shows the feasible selection rule proposed in eq. (4.1). Figure 4.1 shows that choosing the bandwidth parameter by minimising average squared bias reduces discrepancy between the nominal size and the average path coverage[17] of bootstrapped confidence intervals (Panel A vs. C). It also minimises the average squared bias computed against the true series $\{\sigma_t^2\}$ as can be seen in Panel B. In contrast, choosing the bandwidth which minimises the mean square error w.r.t. the GARCH path (Panel D vs C) produces confidence intervals with considerably worse coverage properties.

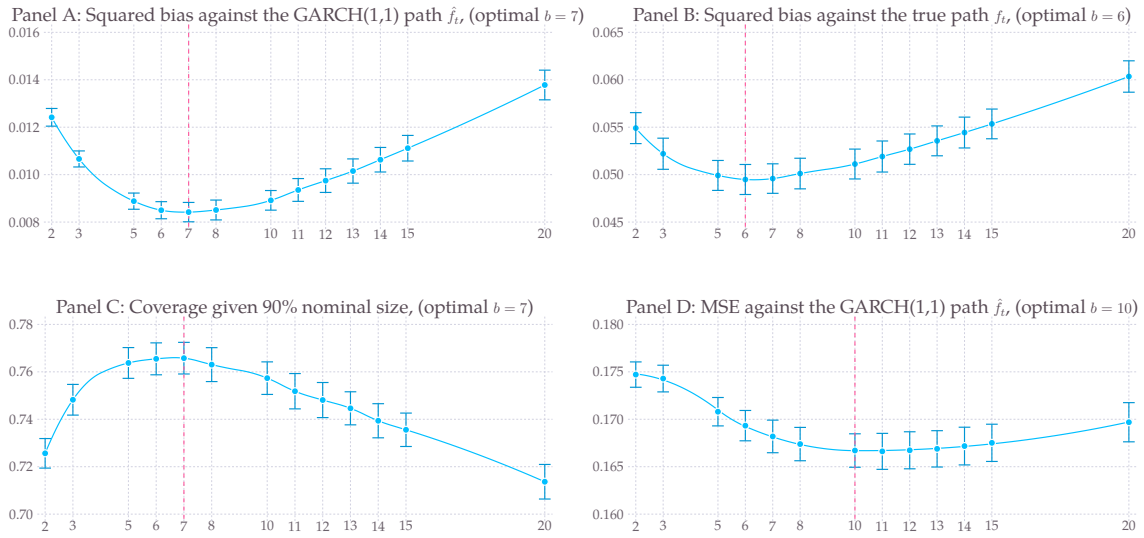
The choice of the optimal bandwidth can be done with fewer bootstrap replications than are required for confidence bands. Estimation of the mean over time requires many

[17] See definition 4.1.

Figure 4.1

Optimal bandwidth selection

In this figure I estimate optimal resampling bandwidth for daily returns on S&P 500 index from the period 1980–2015. To provide simulation results, I first estimate the time-varying variance of these returns with GARCH(1,1). This first filter pass series is then assumed to be the true series $\{\sigma_t^2\}$ and Monte Carlo sample data is created by independently drawing innovations $\eta_t \sim \mathrm{N}(0,1)$ and scaling them with $\{\sigma_t^2\}$. For each such Monte Carlo sample series, the GARCH(1,1) is fitted to obtain $\{\hat{\sigma}_t^2\}$ and the LITE bootstrap method is applied. For each Monte Carlo replication, I compute the average path coverage of a 90% nominal size confidence interval, average squared bias calculated both against $\{\sigma_t^2\}$ and $\{\hat{\sigma}_t^2\}$, as well as mean square error computed against $\{\hat{\sigma}_t^2\}$. All values are reported with their corresponding standard errors captured through error bars. Monte Carlo simulations are performed for various values of the resampling bandwidth parameter, w (shown on the horizontal axis). To accentuate emerging patterns average values of statistics are connected using a quadratic spline. Optimal bandwidths in each panel is drawn in red. In total, 100 Monte Carlo replications were performed for each value of the resampling bandwidth. Bootstrap sample, B, was set at 199.



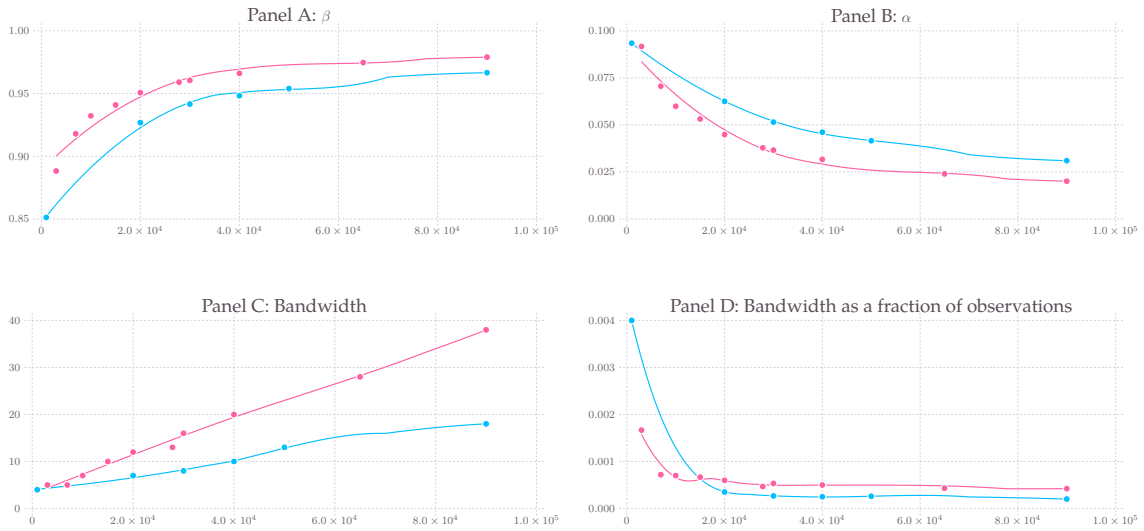
fewer observations than estimating a tail of the distribution. If averaging over time is replaced by taking a median, the number of bootstrap replications can be reduced even further. In real-world situations, I have found that is often sufficient to use as few as $B = 99$ bootstrap replications if the median is employed and $B = 499$ if the average is used.

As the sampling frequency increases, i.e., as we go from monthly, through daily, to hourly (and so forth) data spanning the same period of time, the chosen feasible

Figure 4.2

Optimal bandwidth vs. sample size

This figure presents simulation evidence on behaviour of GARCH parameters (Panel A and B) and the optimal w (Panel C and D) when sampling frequency is being increased. I present results for two different processes, one estimated based on the S&P 500 returns (pink), while the other is based on IBM data (blue). Panel C shows the feasible optimal bandwidth obtained based on the algorithm proposed in this section and is presented in absolute terms. Panel D presents the bandwidth as a fraction of all observations. Horizontal axis denotes the effective number of observations.



bandwidth changes. In Figure 4.2, I present simulation evidence on the behaviour of the resampling bandwidth, w, as the sample size increases in this fashion. Panels A and B show that as we increase the number of observations estimated persistence, β, tends to one, while the learning rate, α, tends to zero. This is in line with results in Nelson (1992); Nelson and Foster (1994), and Jensen and Lange (2010). At the same time, Panel C shows that the size of the optimal bandwidth increases in absolute terms, but decreases as a fraction of the sample length (Panel D). The results I present are for two different processes, one based on the S&P 500 returns and the other on IBM data. Interestingly, given the same estimated persistence, e.g., if $\beta = 0.95$ (for 20,000 and 40,000 observations, respectively), the optimal bandwidth is estimated at roughly the same number of observations (10).

4.7. Empirical application

In this section I demonstrate the usefulness of the LITE bootstrap by evaluating its performance in both empirical and simulation settings. I first bring the method to real-world data. Section 4.7 contains results obtained for S&P 500 returns. Simulation evidence is shown in Section 4.7. In simulations, I set the smoothed (LITE) volatility path obtained for S&P 500 returns as the true path. By doing so, I am able to compare the performance of the LITE confidence bands with a situation where both the true path of volatility and conditional densities are known and it is possible to sample from them infinitely many times. Note that in this way I can also compare the LITE bootstrap with a parametric bootstrap around the GARCH path.

Volatility of S&P 500 returns

I construct confidence intervals for volatility of daily returns on the S&P 500 index. Results for additional indices and stocks are presented in Appendix 4.A. The return series is constructed by taking log differences of daily closing prices in the period 1980–2015 (9040 observations). The total return index was obtained from Datastream. For this data, I estimate the GARCH(1,1) model and provide the LITE confidence intervals as described in Section 4.3. The bandwidth parameter was chosen by minimising average squared bias as described in Section 4.6 and is set at $w = 7$, i.e. for an observation at time t, raw residuals from up to one week earlier and up to one week later are considered as valid replicates. In simulations, the value of this parameter was confirmed to maximise average coverage, see Figure 4.1.

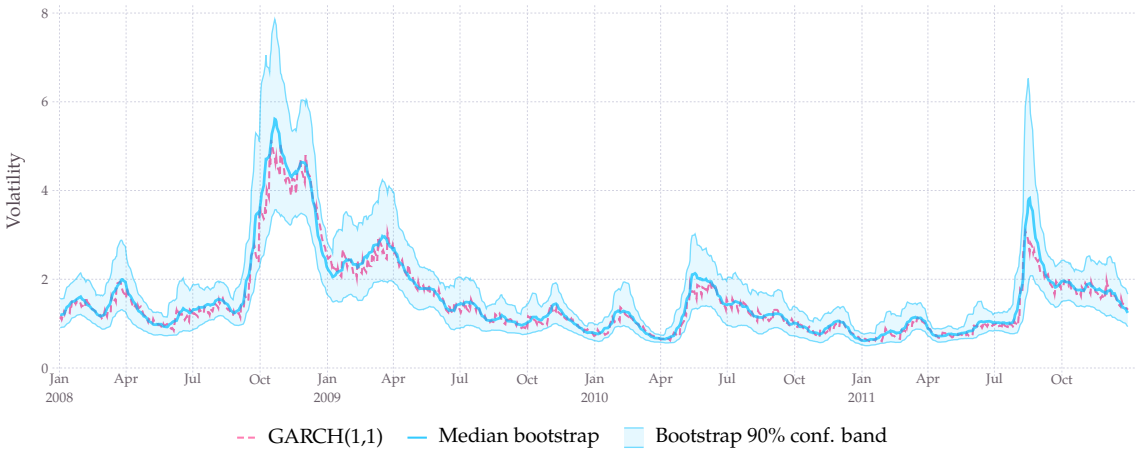
The QMLE estimates for the GARCH(1,1) parameters are $\omega = 0.0145$ (6.5652), $\alpha = 0.0789$ (13.626), and $\beta = 0.910$ (128.6) with values of the t-statistic in parentheses. Figure 4.1 shows the estimated path in a 7-year-long period between 2008 and 2015. The GARCH estimated path is plotted as a dashed red line. As is typical for daily stock returns, the GARCH path exhibits sudden and relatively large changes. At a smaller scale, the GARCH path oscillates around what can be presumed is the true level of

Figure 4.1

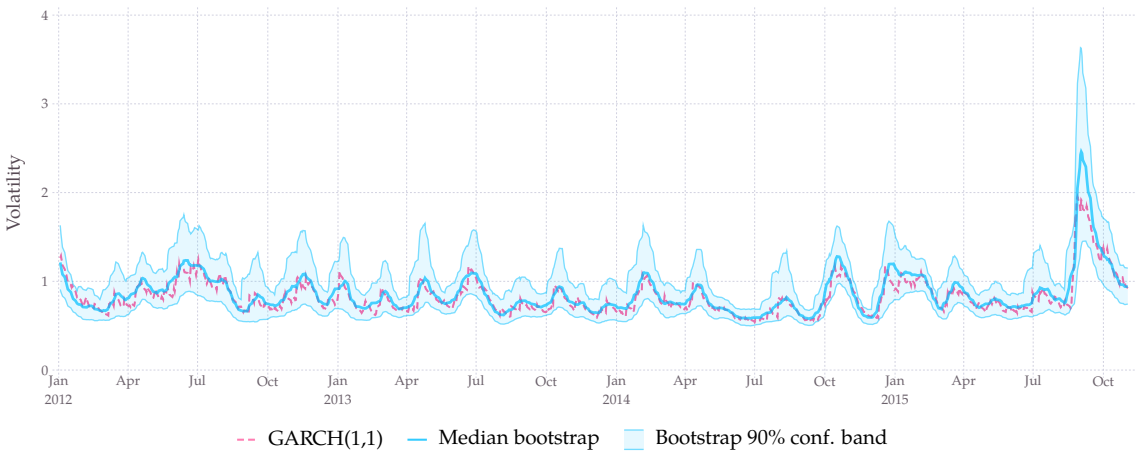
Volatility of S&P 500 returns with confidence bands

Confidence bands for GARCH(1,1) volatility filtered from S&P 500 returns calculated based on closing prices of the total return index. The estimation period is 1980–2015 (9040 observations). For clarity, I only report results for the last 7 years. Panel A contains bands for the period 2008–2012 while Panel B shows the period 2012–2015. Note the change of scale on the vertical axis between the two panels. Bootstrap series were constructed with the lite resampler. The optimal bandwidth of 7 days was determined by minimising average squared bias of the bootstrapped paths which also maximises coverage given the number of observations and the assumption that the original GARCH(1,1) path is true. Red line shows the QMLE estimated path of the time-varying volatility. The blue line shows the median (and average as they coincide) value of volatility across all bootstrapped paths at time t. The light blue region around the median bootstrap path corresponds to a confidence band with 90% nominal size computed using $B = 9999$ bootstrap replications.

Panel A: 2008–2012



Panel B: 2012–2015



volatility given the consistency results in Nelson (1992).

Figure 4.1 also contains 90% confidence bands around the GARCH path. Note that in periods when the volatility suddenly increases—e.g., in October of 2008 and in August of both 2011 and 2015—the bootstrapped confidence band reacts sharply and the median bootstrap path which is denoted as a solid blue line is above the GARCH path. All these cases, however, are relatively short-lived. This suggests that, at the QMLE values of parameters, either the GARCH filter is too slow to adjust or the returns are drawn from a fat-tailed distribution. If the former is true, the median bootstrap path might reflect the true level of volatility more accurately. The confidence bands are for the most part symmetric and the mean and median paths overlap. Only sporadically do the mean and median diverge, which can be better seen in the examples presented in Appendix 4.A. Furthermore, the median bootstrap path offers a smoother path than the GARCH filter. Note how the GARCH path oscillates around the median bootstrap path. Based on the simulation evidence in the remainder of the chapter, the median bootstrapped path could be lessening the impact of the attenuation bias and actually deliver an estimate of the true volatility with a lower average root mean square error.

As can be expected, the span of the confidence bands changes with the value of volatility. This behaviour is driven by the changing signal-to-noise ratio. When volatility is 'high', the amount of information in the data about the conditional variance is lower than when the true volatility is 'low'. This is because a series of high (in absolute values) returns is a strong signal that the volatility has increased. In contrast a series of returns which are close to zero is not as strong a signal that volatility is decreasing. However, the width of the LITE confidence bands is not the same for similar values of volatility. This can be observed both after sharp increases in volatility, i.e., when the median bootstrap path is above the GARCH path, and when the GARCH path is relatively stable, e.g., in late 2011 or in the summer of 2015. This is due to the smoothing behaviour embeded in the LITE procedure.

Simulations

In this section, I explore properties of the LITE confidence bands in a range of simulation experiments. To that end, I take the median bootstrap path obtained in Section 4.7 and assume it to be the true volatility path in Monte Carlo studies. As will become evident in this section, the median bootstrap produces a path characterised by smaller root mean square error. The choice of the GARCH or the median bootstrap path does not qualitatively change the results. The median bootstrap does, however, contain sharper spikes in volatility which are harder for the GARCH filter to match. Results in this section are based on $M = 1000$ Monte Carlo replications and $B = 399$ bootstrap samples.

For each Monte Carlo replication, I create a synthetic time-series of observations by independently drawing a series of standard normal residuals, η_t, and scaling them using the assumed-to-be-true path of volatility as implied by the data-generating process in Section 4.2 which I restate here for convenience:

$$y_t = \varepsilon_t, \tag{4.1}$$

$$\varepsilon_t = \eta_t \sqrt{\hat{\sigma}^2_{t,SPX}}, \qquad \eta_t \sim \mathrm{N}(0,1)\,. \tag{4.2}$$

The true path is shared across all replications and as it is deterministic, the GARCH(1,1) filter is misspecified in the sense described in Section 4.2. Given a Monte Carlo sample series, the GARCH(1,1) filter is used to obtain an estimated path of volatility. This path is then used to obtain 90% confidence bands with the LITE bootstrap along with a median bootstrap path. I keep the resampling bandwidth choice of $w = 7$ days from the previous section.

Furthermore, I also compute a single confidence band based on all (1000 Monte Carlo) GARCH paths which I denote as 'GARCH MC' confidence band. Since the 'GARCH MC' band is obtained in an idealised setting it shows the limits of the filter. I use it as a benchmark for the bands produced by the LITE bootstrap.

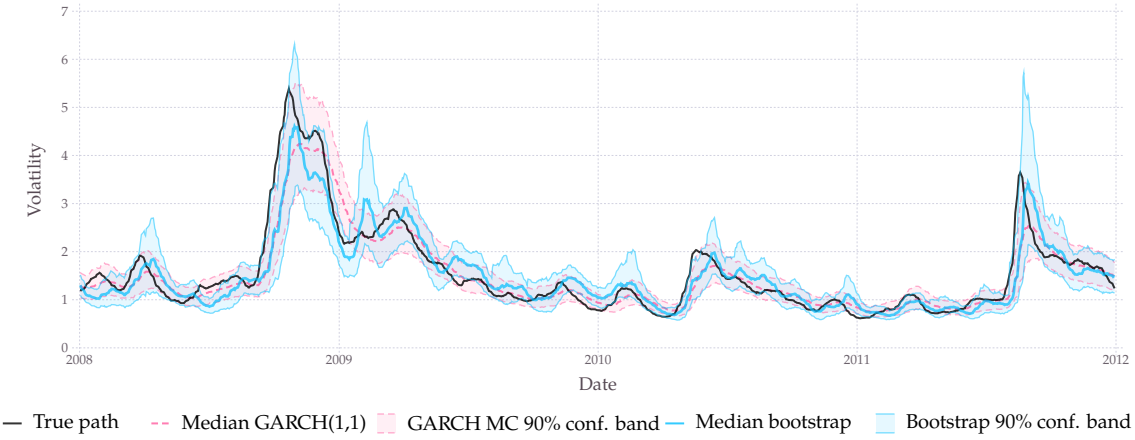
Figure 4.2

Quality of confidence intervals

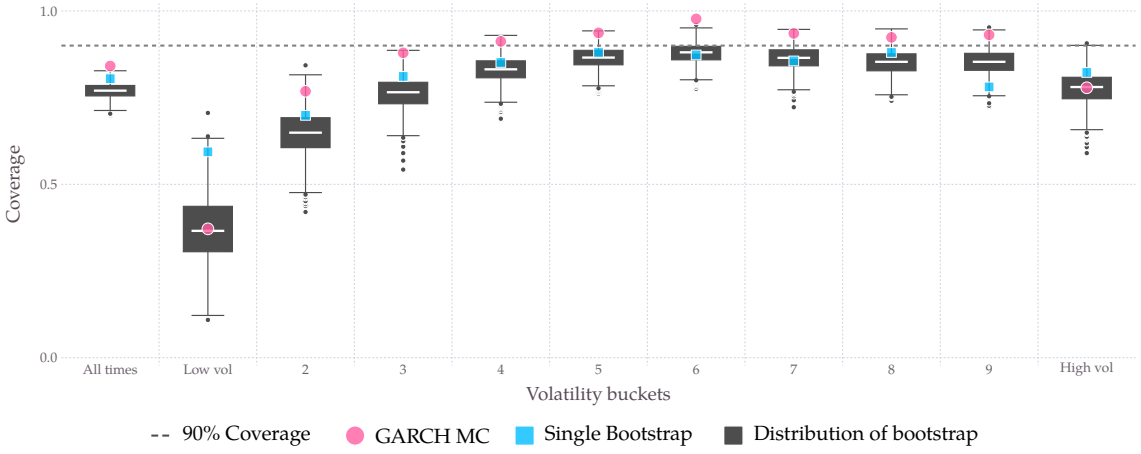
This figure shows the coverage properties of confidence bands obtained with the LITE bootstrap based on a Monte Carlo experiment. All Monte Carlo replications share the same true path of volatility, $\{\sigma_t^2\}$, which is drawn in black in Panel A. Data is generated by independently drawing innovations from a standard normal distribution and scaling them with $\{\sigma_t^2\}$. For each such sample (i) a GARCH(1,1) filter is used to obtain a 'GARCH' path, and (ii) confidence bands are obtained with the LITE bootstrap method (with resampling bandwidth of 7 observations). In total 1000 Monte Carlo replications are performed with $B = 399$ bootstraps. Panel A shows two sets of confidence bands. In blue, is a confidence band produced by the bootstrap for a single Monte Carlo sample. As a benchmark, I also show in red Monte Carlo confidence band based on all 1000 Monte Carlo replications.

Panel B depicts actual coverage of a 90% nominal confidence band for different levels of volatility. As in Panel A, red circles and blue squares show coverage obtained with GARCH(1,1) Monte Carlo and the single bootstrap. Box plots show the distribution of coverage for bootstrapped confidence bands across all Monte Carlo replications.

Panel A: A single bootstrapped confidence interval juxtaposed with Monte Carlo bands



Panel B: Real coverage of a 90% nominal confidence band vs. deciles of true volatility



Quality of confidence intervals

Figure 4.2 contains the first set of results. In Panel A, I plot the GARCH MC band along with bootstrapped bands from a single Monte Carlo replication. Even at a daily frequency, both bands are close in terms of both how tight they are at most times and how well they track the true path.

Panel B summarises coverage for different deciles of the true volatility. I use the notion of the average path coverage from Definition 4.1. Unconditional average path coverage for LITE bands is distributed around 77%. Bootstrap bands for the single Monte Carlo replication in Panel A have an average path coverage of 79%. This is slightly lower than the coverage of the benchmark, GARCH MC, which is at 84%. The fact that average path coverage is below the nominal level is not surprising. lite is a non-parametric percentile bootstrap procedure and these are known to produce shorter confidence intervals (Efron, 2003). The bands can be potentially improved upon with either the BC_a or the ABC method of computing percentiles (Efron and Tibshirani, 1994; DiCiccio and Efron, 1996). However, these improvements add another layer of complexity. In particular, at this stage, it is not obvious how to estimate additional shape parameters these methods require.

Interestingly, coverage of both the LITE and the GARCH MC bands is not uniform and depends heavily on the value of volatility itself. When volatility is at its very lowest, both types of bands have on average only 37% path coverage. This is due to the fact that for these values paths are over-estimating the true volatility. Furthermore, coverage of the bands is not stable in this region which makes it difficult to correct for the bias even though it is evident that bias correction based on the estimated value of volatility is possible. The fact that the GARCH MC band also suffers from this issue is comforting as the deficiency clearly stems from the filter and not from the LITE procedure. In fact, Nelson and Foster (1994, p. 22) foresee this outcome and attribute it to the inability of the GARCH to match the true 'variance of the variance' everywhere in the state space. It also suggests that a more robust choice of filter can improve results.

In regimes with higher volatility, path coverage quickly increases towards the nominal size of 90% and is on average 86% for medium and medium-high volatilities. Note that LITE bands have lower than nominal coverage in general. Average coverage of the benchmark—GARCH MC—bands is below (above) the nominal level for low and high (medium) levels of volatility.

Bias and RMSE of median bootstrap paths

As mentioned before, the LITE procedure can act as a smoother with the median bootstrap path representing the smoothed estimator of $\{\hat{\sigma}_t^2\}$. Figure 4.3 provides evidence for this claim. In Panel A, I again show the GARCH MC bands as a benchmark. In addition, the plot contains Monte Carlo bands based on the 1000 median bootstrap paths. The two sets of bands overlap almost fully. The only exceptions are, again, periods where the true volatility suddenly increases. Median bootstraps track these changes much better and as a consequence the confidence bands are shifted towards the true value. This can be seen in Panel B of Figure 4.3 which contains distributions of both bias and root mean square error produced by the GARCH and the median paths. As stated before, the root mean-square error for the median bootstrap paths is lower than for GARCH paths. This is also a typical case of trading in bias for a reduction in variance. In this case, this reduction is substantial.

Suboptimal choice of resampling bandwidth

The resampling bandwidth has sizeable effects on coverage properties of confidence bands produced by the LITE bootstrap. In Figure 4.4, I compare the situation in which the bandwidth was either set lower or higher than the optimal level based on the rule in Section 4.6. Panel A of this figure again compares the GARCH MC[18] bands with the LITE bands from a single Monte Carlo replication. When bandwidth is set too low,

[18] All results for GARCH MC, including bands and coverage are exactly the same as in Figure 4.2 and Figure 4.3.

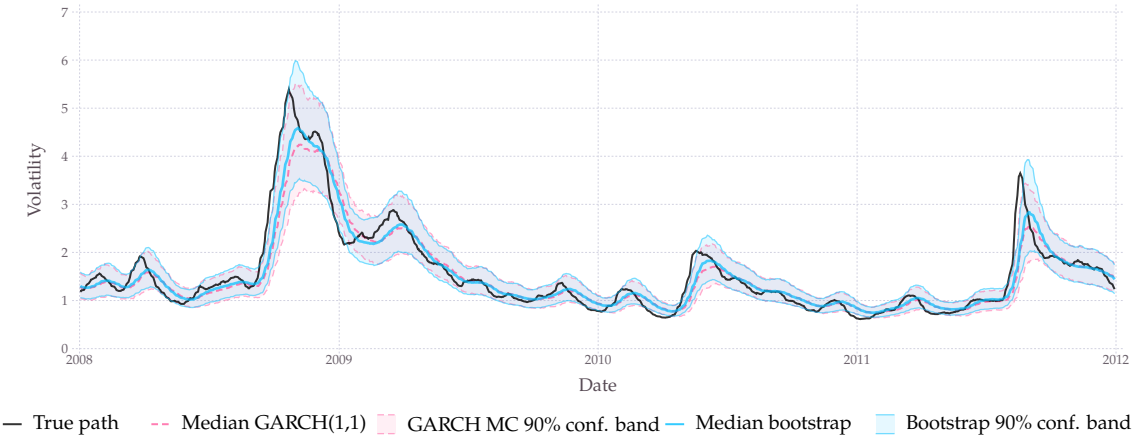
Figure 4.3

Bias and RMSE of median bootstrap paths

In this figure I provide evidence for consistency of paths estimated using the LITE bootstrap. Panel A shows results in the time dimension. A path of volatility which is depicted in yellow is considered to be true in a Monte Carlo setting. Data is generated by drawing independently innovations from a standard normal distribution and scaling them with the yellow path. For each such sample (i) a GARCH(1,1) filter is used to obtain a 'GARCH' path, and (ii) confidence bands are obtained with the LITE bootstrap method (setting the resampling bandwidth at 7 observations). In total 1000 Monte Carlo replications are performed with $B = 399$ bootstraps. Panel A shows two sets of confidence bands. In blue, is a Monte Carlo confidence band produced by all 1000 estimated median bootstrap paths. As a benchmark, similarly constructed confidence bands out of the GARCH paths are shown in red.

Panel B shows the distribution of average bias and average root mean square error (RMSE) across all Monte Carlo simulations. Average here is computer over the time dimension for each replication.

Panel A: Monte Carlo confidence bands for both GARCH and median bootstrap paths



Panel B: Bias and root mean square error for both GARCH and median bootstrap paths

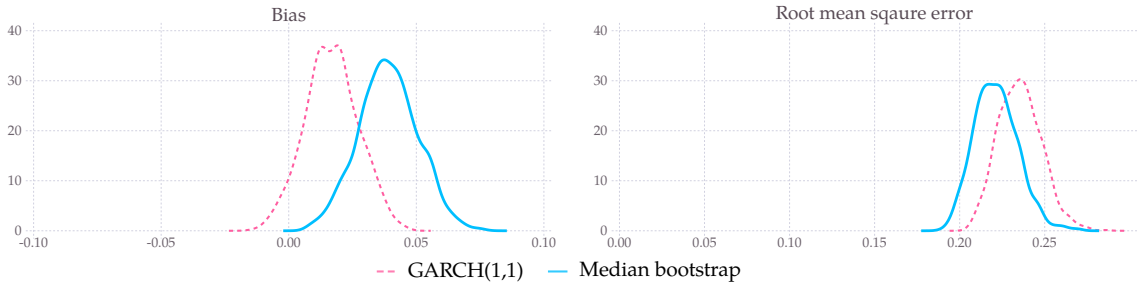
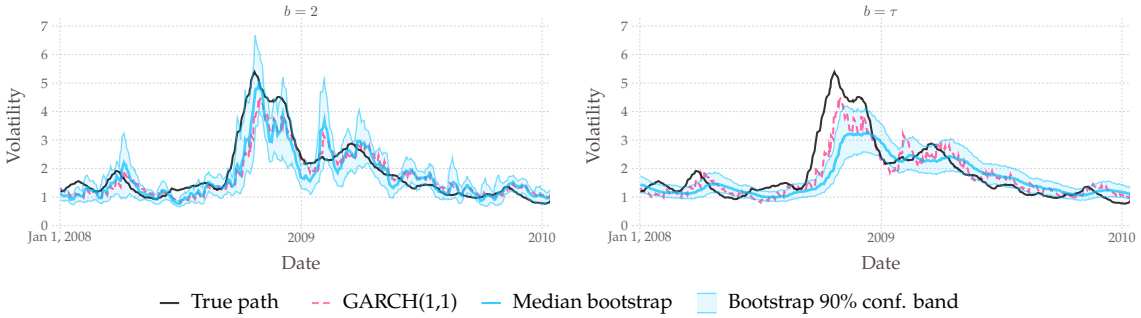


Figure 4.4

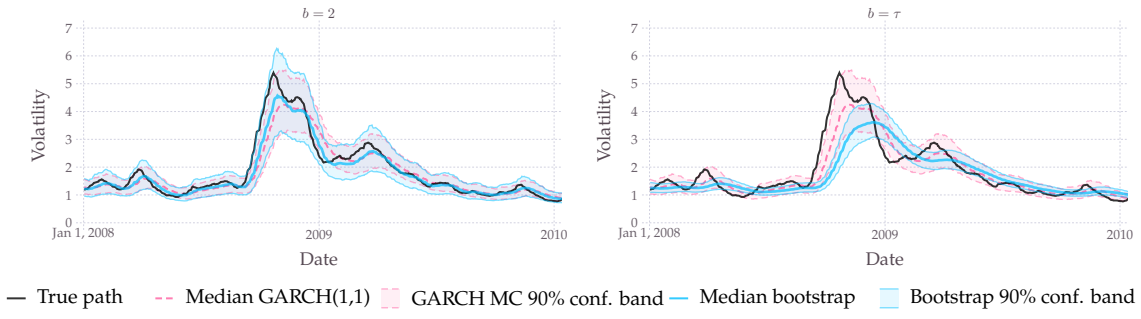
Suboptimal choice of resampling bandwidth

In this figure I show the coverage properties of confidence bands obtained using the LITE bootstrap procedure when resampling bandwidth is not chosen by minimising average squared bias. In each panel, the figure on the left sets the resampling bandwidth at 2 days which is smaller than the optimal 7 days. In figures on the right, the bandwidth was set to the number of observations, this recovers the *i.i.d.* bootstrap of Efron (1979). Panels A, B, and C replicate results from Figure 4.2 and Figure 4.3 which offer detailed descriptions.

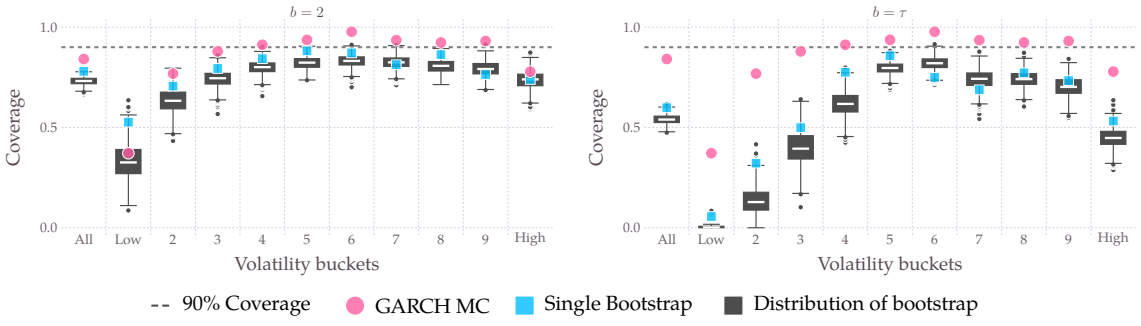
Panel A: Bootstrapped confidence interval



Panel B: Monte Carlo results for median bootstrap and GARCH paths



Panel C: Real coverage of a 90% nominal confidence band vs. deciles of true volatility



the attenuation bias[19] is strengthened. The median bootstrap is even less persistent than the GARCH path. At the same time, confidence bands are narrower which has a universally negative effect on path coverage shown in Panel C.

If bandwidth is set too high, as is the case in the Efron's rule, the bootstrapped bands provide poor coverage at approximately 60%. As is clear from Panel A, this is because the bootstrap paths are over-smoothed and are significantly lagging behind the GARCH path. These results are not surprising given the discussion Section 4.6 and only serve to further motivate the feasible bandwidth selection which I propose therein.

Finally, the way in which the bootstrap bands lag behind the GARCH path in Panel A is similar to how the Monte Carlo confidence bands lag behind the true volatility path in Panel B. Further, note that the Monte Carlo experiment can be actually seen as a semi-parametric bootstrap with *i.i.d.* resampling. Based on these two observations and all the simulation results presented so far, I conclude, that the bootstrap confidence bands obtained in the empirical part of the chapter, Section 4.7, are likely to also have coverage close to the nominal level of 90%.

Convergence

Performance of the LITE bootstrap should improve if the sampling frequency is increased. In this section, I explore coverage properties when the process is sampled every three hours as opposed to daily. To that end, I fit a high-degree B-spline curve to the median bootstrap volatility path for S&P 500 returns which was previously set as the 'true path'. This has several advantages. First of all, I avoid any micro-structure effects and diurnal patterns in volatility which could affect results based on empirical intra-day data. Secondly, in this way results are directly comparable to simulations discussed above. An increase in the sampling frequency means that the volatility process becomes more persistent.

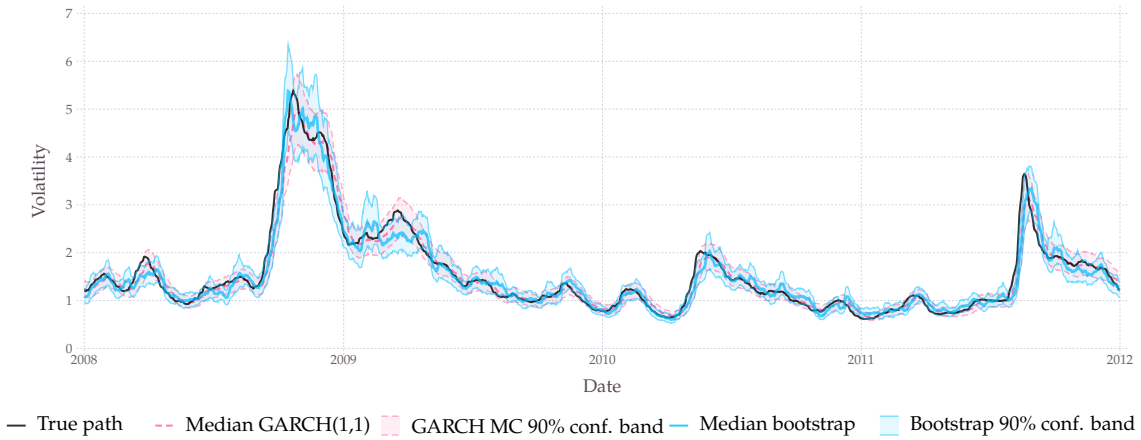
[19] Attenuation bias is a common issue in volatility modelling wherein paths estimated with real-world data are characterised by smaller persistence than the true volatility process.

Figure 4.5

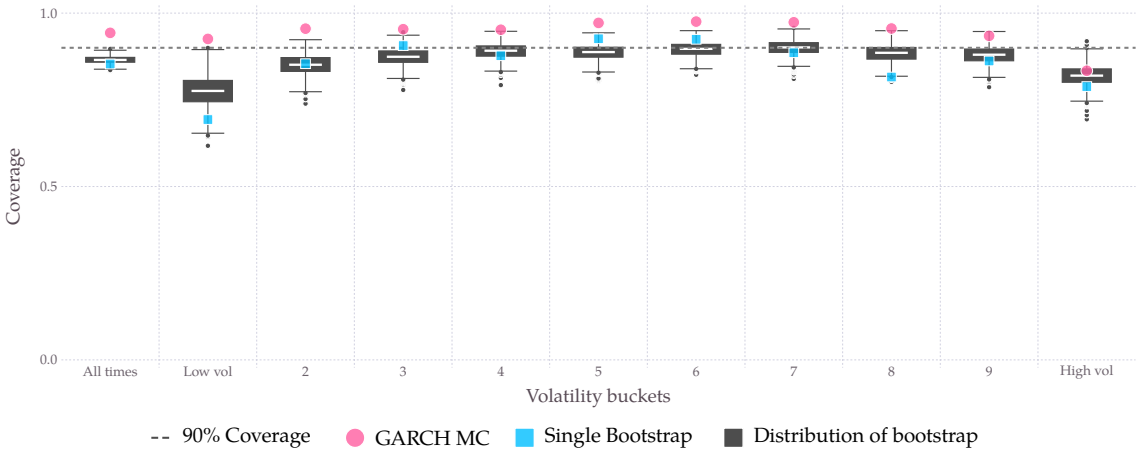
Convergence

In this figure I show the coverage properties of LITE confidence bands in the case where the true volatility process is sampled eight times more frequently (equivalent of 3-hour observations) than in the previous examples. The optimal bandwidth was set at $w = 24$ periods which translates to three days. The single bootstrap band was computed with $B = 4999$ bootstrap replications. A more detailed description is given for Figure 4.2 which this figure replicates.

Panel A: Bootstrapped confidence interval



Panel B: Real coverage of a 90% nominal confidence band vs. deciles of true volatility



In this case, the optimal bandwidth is chosen to be 24 observations based on the selection rule from Section 4.6. Compared to daily returns, the moving-window increases from 15 to 49 observations. However, the fraction of observations considered as valid replicates shrinks by half, from 15 days to only 7 days.

Figure 4.5 summarises results of this analysis. First of all, Panel A shows that the LITE bootstrap intervals are remarkably similar to the GARCH MC bands. The median bootstrap and GARCH MC paths track the true value precisely which shows convergence of the GARCH filter in line with the theoretical results in Nelson (1992) and Nelson and Foster (1994). Furthermore, LITE bands are just as tight as the ideal, GARCH MC bands.

As sampling frequency is increased, the average path coverage of the LITE confidence bands converges to the nominal level. At a 3-hour frequency, unconditional path coverage is on average 86% based on the results in Panel B. The increase is mostly driven by vast improvements of coverage in extremely low and high volatility regimes. It is worth noting, that the confidence bands appear to be accurate on both sides. Frequencies with which the true volatility falls outside of the confidence band on either side appear to be equal and close to the nominal value of 5% For medium volatility regimes LITE bands have an on average 89% path coverage. Interestingly, LITE intervals remain permissive across the board, i.e. they all approach the nominal level from below. Conversely, GARCH MC bands are too conservative and on average cover the true value 95% of the time.

Another improvement from the increase in sampling frequency (or persistence) is seen in stability of the confidence intervals as evidenced by the smaller box-plots in Panel B of Figure 4.2 if compared to Figure 4.2. In other words, improvement in stability means that the interquartile range and standard deviation of the path coverage statistic decrease. This is of great significance even without the simultaneous improvement in the average coverage level (Efron, 2003). Stability allows to more reliably bias correct the confidence bands by stretching them depending on the value of the estimated volatility (or by obtaining them using a higher nominal level).

4.8. Conclusions

This chapter introduces a novel bootstrap procedure which allows me to compute in-sample confidence bands for misspecified GARCH filters. The bands account for two sources of uncertainty, including the parameter uncertainty which stems from the fact that 'pseudo-true' parameters are unknown and merely estimated. More importantly, the LITE procedure also accounts for the filtering uncertainty which comes into play when the GARCH filter is not the true data-generating process. To my knowledge, this is the only procedure which accounts for both sources of uncertainty.

In simulations, average path coverage of the LITE confidence bands is close to the nominal size. At any fixed sampling frequency, the path coverage is state-dependent and, in particular, coverage is lower when the true volatility is very low. I show, however, that this deficiency stems from the filter and can be potentially resolved by choosing a different GARCH model. In general, the LITE bands behave very similarly to the Monte Carlo confidence bands which could be constructed had the true volatility path be known. Furthermore, LITE can be used to produced a smoothed estimate of the true volatility paths by extracting the median bootstrap path. Based on simulation evidence, the smoothed volatility paths have lower root mean square error. Furthermore, the 'smoother' visibly improves fit after sudden and sizeable increases in volatility.

Finally, although the chapter focuses on the family of GARCH filters, the method itself can be potentially applied to other observation-driven or parameter-driven methods as long as consistency of these filters can be shown or assumed.

4.A: Additional empirical results

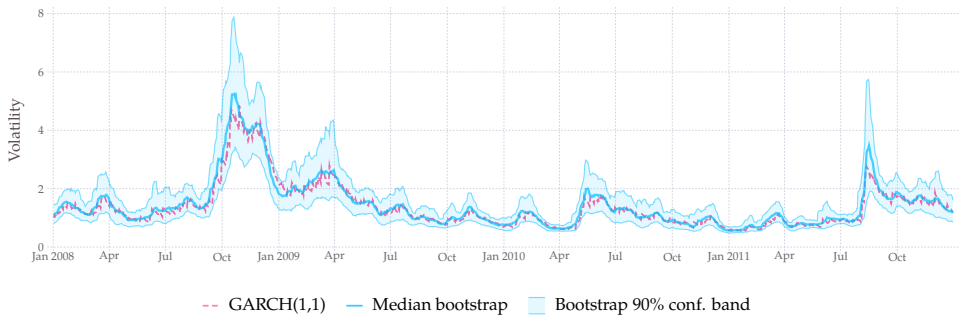
Here, I provide results for additional real-world samples of daily returns. LITE confidence bands were computed for Dow Jones Industrial Average and MSCI World indices in Figure 4.A.1. Figure 4.A.2 contains results for selected U.S. equities: IBM, Procter & Gamble, and 3M. Apart from MSCI World, which starts in 2003, sample periods were set at 1980–2015. For each series an optimal bandwidth was selected according to the rule in Section 4.6.

Figure 4.A.1

Confidence bands for volatility: world indices

This figure shows GARCH(1,1) estimated volatility for a selection of world indices along with LITE confidence bands. Although the sample period is longer, I only report results for 2008–2012. See Figure 4.1 for detailed description of all paths. LITE confidence intervals were computed with $B = 1999$ replications.

Panel A: Dow Jones Industrial Average, bandwidth is set to 5 days



Panel B: MSCI World, bandwidth is set to 5 days

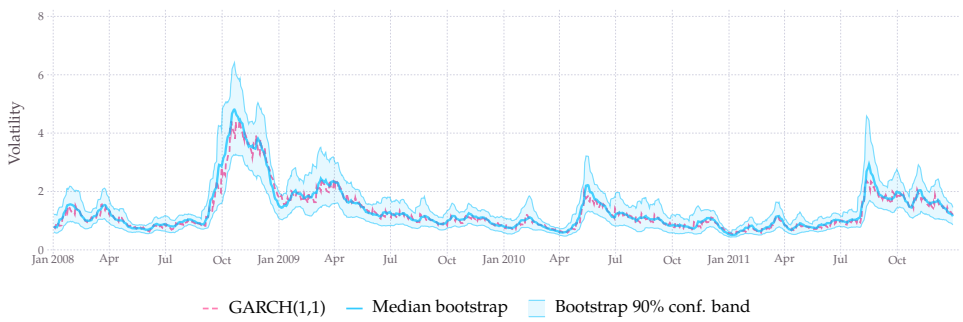
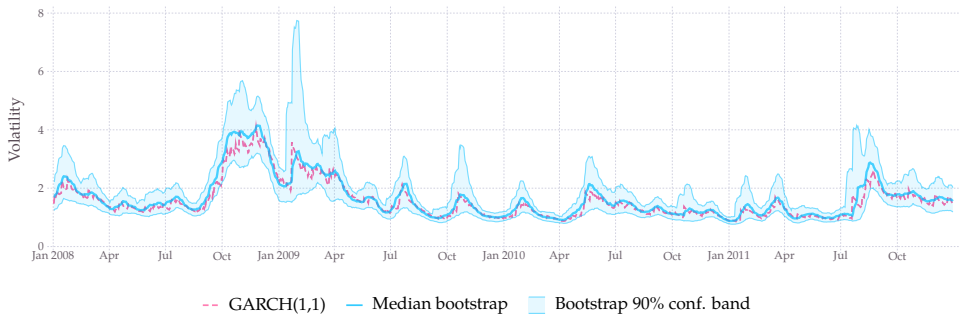


Figure 4.A.2

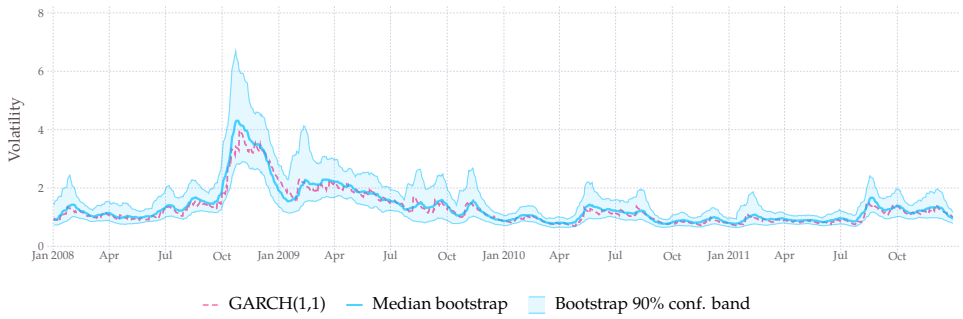
Confidence bands for volatility: U.S. stocks

This figure shows GARCH(1,1) estimated volatility for a selection of U.S. stocks along with LITE confidence bands. Although the sample period is longer, I only report results for 2008–2012. See Figure 4.1 for detailed description of all paths. LITE confidence intervals were computed with $B = 1999$ replications.

Panel A: IBM, bandwidth is set to 7 days



Panel B: Procter & Gamble, bandwidth is set to 9 days



Panel C: 3M, bandwidth is set to 12 days